

Análisis de sentimientos a nivel de aspecto usando ontologías y aprendizaje automático

Aspect-based sentiment analysis using ontologies and machine learning

Carlos Henríquez^(1,2), Ferran Pla⁽³⁾, Lluís-F. Hurtado⁽³⁾, Jaime Guzmán⁽¹⁾

⁽¹⁾Universidad Nacional de Colombia

Cl. 59a #63-20, Medellín, Antioquia, Colombia

jaguzman@unal.edu.co

⁽²⁾Universidad Autónoma del Caribe

Cl. 90 #46-112, Barranquilla, Atlántico, Colombia

chenriquez@uac.edu.co

⁽³⁾Universitat Politècnica de València

Camí de Vera, s/n, 46022 Valencia, España

{fpla, lhurtado}@dsic.upv.es

Resumen: En este artículo se presenta un sistema de análisis de sentimientos a nivel de aspecto que permite extraer automáticamente las características de una opinión y determinar la polaridad asociada. El sistema propuesto está basado en un modelo que utiliza ontologías de dominio para la detección de los aspectos y un clasificador basado en Máquinas de Soporte Vectorial para la asignación de la polaridad a los aspectos detectados. El trabajo experimental se ha realizado utilizando el conjunto de datos desarrollado para la Tarea 5, Sentence-level ABSA en SemEval 2016 para el español. El sistema propuesto ha obtenido un 73.07 en F_1 en la extracción de aspectos (*slot2*) y un 46.24 de F_1 en la subtask conjunta de categorización y extracción de aspectos (*slot1,2*) utilizando una aproximación basada en ontologías. Para la subtask de clasificación de sentimientos (*slot3*) se ha obtenido una *Accuracy* de 84.79 % utilizando una aproximación basada en el uso de Máquinas de Soporte Vectorial y lexicones de polaridad. Estos valores superan los mejores resultados obtenidos en SemEval.

Palabras clave: Análisis de sentimientos a nivel de aspecto, ontologías, máquinas de soporte vectorial

Abstract: In this paper, we present an aspect-based sentiment analysis system that allows to automatically extract the characteristics of an opinion and to determine their associated polarity. The proposed system is based on a model that uses domain ontologies for the detection of aspects and a classifier based on the Support Vector Machines formalism for assigning the polarity to the detected aspects. The experimental work was conducted using the dataset developed for Task 5, Sentence-level ABSA in SemEval 2016 for Spanish. The proposed system has obtained a 73.07 in F_1 in the aspect extraction subtask (*slot2*) and a 46.24 of F_1 in the categorization and aspect extraction subtask (*slot1,2*) using an ontology-based approach. For the sentiment classification subtask (*slot3*) an 84.79 % in terms of *Accuracy* has been obtained using an approach based on Support Vector Machines and polarity lexicons. These results are better than those reported in SemEval.

Keywords: Aspect-based sentiment analysis, ontologies, support vector machines

1 Introducción

Hoy en día en Internet se producen millones de datos debido a la utilización masiva de las redes sociales, servicios de mensajería, blogs,

wikis, comercio electrónico, entre otros. Toda esta gran cantidad de datos es atractiva para diferentes estamentos comerciales, industriales y académicos, pero la extracción y su

respectivo procesamiento, hace que esta tarea sea muy compleja y difícil si se hace de forma manual. Sumado a esto, los usuarios participan activamente en Internet dejando sus propios comentarios, opiniones y reseñas sobre todo tipo de temas.

Debido a esto, los investigadores vienen trabajando desde hace varias décadas en sistemas que permiten analizar gran cantidad de textos de forma automática usando técnicas de procesamiento de lenguaje natural (PLN) y minería de datos, entre otras.

El análisis de sentimientos (AS) es una área del PLN cuyo objetivo es analizar las opiniones, sentimientos, valoraciones, actitudes y emociones de las personas hacia determinadas entidades como productos, servicios, organizaciones, individuos, problemas, sucesos, temas y sus atributos (Liu, 2012). Es decir, extraer una opinión, analizarla y determinar su polaridad (positiva, negativa o neutra).

La gran mayoría de los enfoques para el AS detectan sentimientos a nivel general en una frase, un párrafo o un texto completo (Steinberger, Brychcín, y Konkol, 2014). Este tipo de análisis, conocido como AS a nivel de documento o global, busca clasificar el sentimiento de todo un documento como positivo o negativo (Pang y Lee, 2008). Otros enfoques intentan obtener la polaridad a nivel de frase o a nivel de aspectos. El nivel de frase clasifica el sentimiento expresado en cada oración y el de aspectos lo clasifica con respecto a las características específicas de las entidades encontradas (Medhat, Hassan, y Korashy, 2014). Los dos primeros enfoques resultan a veces incompletos ante la realidad de las empresas u organizaciones que quieren saber en detalle el comportamiento de sus productos (Xianghua et al., 2013). Según (Liu, 2015) el AS a nivel de documento y frase resulta insuficiente para descubrir las preferencias de los usuarios.

El Análisis de Sentimientos a nivel de aspectos (aspect-based sentiment analysis) o Análisis de Sentimientos basado en características (feature-based sentiment analysis) tiene como objetivo identificar las propiedades o características de una entidad y determinar la polaridad expresada de cada aspecto de esa entidad (Hu y Liu, 2004; Liu, Hu, y Cheng, 2005).

En este artículo se presenta un sistema de análisis de sentimientos a nivel de aspecto que

combina ontologías para extraer los aspectos de una entidad y un sistema de aprendizaje automático basado en Máquinas de Soporte Vectorial (SVM) para determinar su polaridad.

El resto del artículo está organizado de la siguiente manera. En la Sección 2 se abordan los antecedentes y trabajos relacionados. La Sección 3 describe el sistema propuesto. La Sección 4 muestra los experimentos realizados y los resultados obtenidos, y finalmente en la Sección 5 se presentan algunas conclusiones y trabajos futuros.

2 *Antecedentes y trabajos relacionados*

Para la construcción de un sistema de AS a nivel de aspecto se debe iniciar con la extracción de los aspectos de la opinión. En la literatura existen diferentes enfoques para esta tarea. En (Wang, Lu, y Zhai, 2010) se utiliza una lista ya predeterminada de aspectos. En (Zhang, Xu, y Wan, 2012) se usa conteo de nombres y frases para calcular su frecuencia dentro de un documento. (Qiu et al., 2011) aprovechan las relaciones entre sentimiento y aspectos. (Marcheggiani et al., 2014) se basan en modelos de aprendizaje supervisado. (Xianghua et al., 2013) utilizan modelos estadísticos LDA (Latent Dirichlet Allocation) y (Poria et al., 2016) mejoran estos modelos estadísticos usando similitud semántica.

De todos los enfoques anteriores, la gran mayoría no tiene en cuenta el significado de las palabras que representan a los aspectos. Éstos son considerados simples “etiquetas” que no son situadas en el contexto de la opinión ni en el dominio de la entidad a la cual se está refiriendo. Sin embargo, el enfoque presentado en este trabajo sí tiene en cuenta el significado de los aspectos y utiliza para su extracción ontologías de dominio. Las ontologías consisten en especificaciones formales y explícitas que representan los conceptos de un determinado dominio y sus relaciones, es decir, son un modelo abstracto de un dominio, donde los conceptos utilizados están claramente definidos (Studer, Benjamins, y Fensel, 1998). En la literatura se han usado las ontologías para análisis de sentimiento, entre otros trabajos, en (Peñalver-Martínez et al., 2014), (Cadilhac, Benamara, y Aussenac-Gilles, 2010) y (Kontopoulos et al., 2013). Una comparación de cómo se utilizan se encuentra en (Henríquez y Guzmán, 2016).

A partir del aspecto extraído, el siguiente paso consiste en determinar su polaridad, también conocida como clasificación de sentimiento (Henríquez Miranda, Guzmán, y Salcedo, 2016). Para lograr lo anterior, se distinguen dos enfoques principales: las técnicas basadas en aprendizaje automático (AA) y las basadas en léxico (LEX) (Medhat, Hassan, y Korashy, 2014). Encontramos en la literatura trabajos relacionados directamente con nuestra aproximación. Por ejemplo, en (De Freitas y Vieira, 2013) se realiza un análisis guiado por ontologías en el dominio de cine y hoteles en portugués; (Steinberger, Brychcín, y Konkol, 2014) presentan un enfoque supervisado en opiniones de restaurantes en checo; (Manek, Shenoy, y Mohan, 2016) proponen un sistema en inglés basado en el índice GINI para comentarios en cine; (Marcheggiani et al., 2014) proponen un conjunto de modelos basados en campos aleatorios condicionales para las reseñas de hoteles en inglés; (Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Martínez-Cámara, E., y Ureña-López, 2016) presentan un enfoque no supervisado empleando un método basado en léxico que combina diferentes recursos lingüísticos sobre un conjunto de datos de entrenamiento en inglés sobre los dominios de restaurantes y portátiles.

3 Descripción del sistema

En la Figura 1 se muestra un esquema del sistema propuesto. El sistema consta básicamente de tres módulos: preprocesamiento, extracción de aspectos y clasificación de sentimientos.

3.1 Preprocesamiento

El sistema recibe como entrada un documento D que contiene una o varias opciones acerca de una entidad. En el siguiente paso se segmenta la opinión en oraciones y estas oraciones en palabras con sus correspondientes lemas. En este proceso se ha usado *Freeling* (Padró y Stanilovsky, 2012) para obtener la lematización y el etiquetador morfosintáctico de los textos considerados. Además, se ha realizado un proceso de normalización del vocabulario para corregir algunos términos informales usados en las redes sociales y también para corregir errores tipográficos. En concreto, se han eliminado signos de puntuación, se han eliminado algunas *stopwords*, por ejemplo: *el*, *la*, *lo*, *su*, ..., y se ha usado un dic-



Figura 1: Arquitectura del Sistema

cionario en español para corregir algunas palabras y terminaciones usuales, por ejemplo, cambiar la terminación *ion* por *ión*, etc.

3.2 Extracción de aspectos

A partir del texto etiquetado, los aspectos son extraídos utilizando dos procesos. El primero utiliza una ontología de dominio la cual describe el vocabulario relacionado con un dominio específico (hoteles, cine, restaurantes, ...). El segundo proceso utiliza similitud semántica usando una base de datos léxica (Meng, Huang, y Gu, 2013) que permite encontrar posibles aspectos relacionados con aquellos conceptos que no se extrajeron en el proceso anterior. La salida final del proceso será una lista de aspectos $L(A)$.

Para la primera parte se debe disponer de una ontología de dominio en el lenguaje que se vaya a manejar. En este trabajo se ha utilizado la ontología “Hontology” (Chaves, Freitas, y Vieira, 2012) para analizar comentarios de restaurantes en español. Los sustantivos de la opinión se buscan en la ontología y los encontrados en ella se marcan como aspectos.

Para la segunda parte, se toman los sustantivos no encontrados en la ontología y se calcula una similitud semántica con los conceptos de la ontología. El cálculo de similitud semántica se basa en el algoritmo de camino

```

<sentence id="xxx">
  <text>Buen servicio, ambiente Acogedor y tranquilo, comida bien.</text>
  <Opinions>
    <Opinion target="servicio" category="SERVICE#GENERAL" polarity="positive" from="5" to="13"/>
    <Opinion target="ambiente" category="AMBIENCE#GENERAL" polarity="positive" from="15" to="23"/>
    <Opinion target="comida" category="FOOD#QUALITY" polarity="positive" from="47" to="53"/>
  </Opinions>
</sentence>

```

Figura 2: Ejemplo de una opinión y la anotación de los diferentes slots (“category” corresponde al *slot1*, “target” corresponde al *slot2* y “polarity” al *slot3*)

más corto, donde las palabras que se comparan son nodos en un árbol de dominio en el cual los nodos hijos tienen una relación ‘es un’ con los padres. Por ejemplo, ‘carro es un vehículo’, en esta relación, vehículo es padre de carro.

Al momento de determinar si un sustantivo está directamente relacionado con un elemento de una ontología, se calcula la similitud entre ellos y se valida si la puntuación obtenida es mayor o igual que un umbral definido experimentalmente a partir del conjunto de entrenamiento.

Para el cálculo de la similitud se utilizó Wordnet en español disponible en MCR (Multilingual Central Repository) (Gonzalez-Agirre y Rigau, 2013).

3.3 Clasificación de sentimientos

Una de las dificultades de la tarea consiste en, una vez detectado el aspecto, definir qué contexto se le asigna para poder establecer su polaridad. Para la detección de la polaridad a nivel de aspecto, se ha utilizado una aproximación ya utilizada para el dominio de Twitter y presentada en (Pla y Hurtado, 2014; Hurtado y Pla, 2014). En concreto, en este trabajo se propone una aproximación que consiste en determinar el contexto de cada aspecto a través de una ventana fija definida a la izquierda y derecha del aspecto. La longitud de la ventana se ha determinado experimentalmente mediante una validación cruzada utilizando el conjunto de entrenamiento proporcionado. El valor máximo de la ventana considerado es de 6 palabras a izquierda y derecha del aspecto. Para entrenar nuestro sistema, se ha considerado el conjunto de entrenamiento únicamente, se han determinado los segmentos para cada aspecto y se ha entrenado el clasificador. La misma segmentación utilizada para el entrenamiento se ha aplicado al conjunto de test.

Como clasificador se han utilizado Máquinas de Soporte Vectorial por su capacidad para manejar con éxito grandes cantidades

de características. En concreto usamos dos librerías (*LibSVM* y *LibLinear*) que han demostrado ser eficientes implementaciones de SVM que igualan el estado del arte. El software se ha desarrollado en *Python* y para acceder a las librerías de SVM se ha utilizado el toolkit *scikit-learn*. Los parámetros de los clasificadores se han determinado en la fase de ajuste de parámetros usando una validación cruzada de 10 iteraciones (10-fold cross-validation).

Se han explorado otras aproximaciones de aprendizaje automático para desarrollar el clasificador. En particular, un sistema basado en el uso de redes neuronales convolucionales y recurrentes (Zhou, Wu, y Tang, 2002; Lecun, Bengio, y Hinton, 2015) y en la combinación de “embeddings” de palabras (Mikolov et al., 2013b; Mikolov et al., 2013a). Aunque este sistema ha obtenido resultados prometedores para tareas de SA en inglés y en árabe (González, Pla, y Hurtado, 2017), los resultados obtenidos hasta el momento para la tarea que se presenta en este trabajo son ligeramente inferiores a los alcanzados mediante el sistema basado en SVM.

4 Experimentación y resultados

Para validar el sistema propuesto se realizó una serie de experimentos utilizando el corpus de la tarea 5 de la edición de 2016 de SemEval (International Workshop on Semantic Evaluation). Específicamente, se abordó la subtarea 1 (SB1) en el dominio de restaurantes para el español (Pontiki et al., 2016).

La subtarea SB1, a su vez, está dividida en 3 subtareas, denominadas *slots*. El *slot1* consiste en detectar la categoría-aspecto de una opinión. Cada categoría está compuesta por un par entidad (E), atributo (A) representado como E#A. Se proporciona una lista de un total de 12 categorías (p.e. RESTAURANT#GENERAL, RESTAURANT#PRICES, FOOD#QUALITY). Es posible asociar más de una categoría

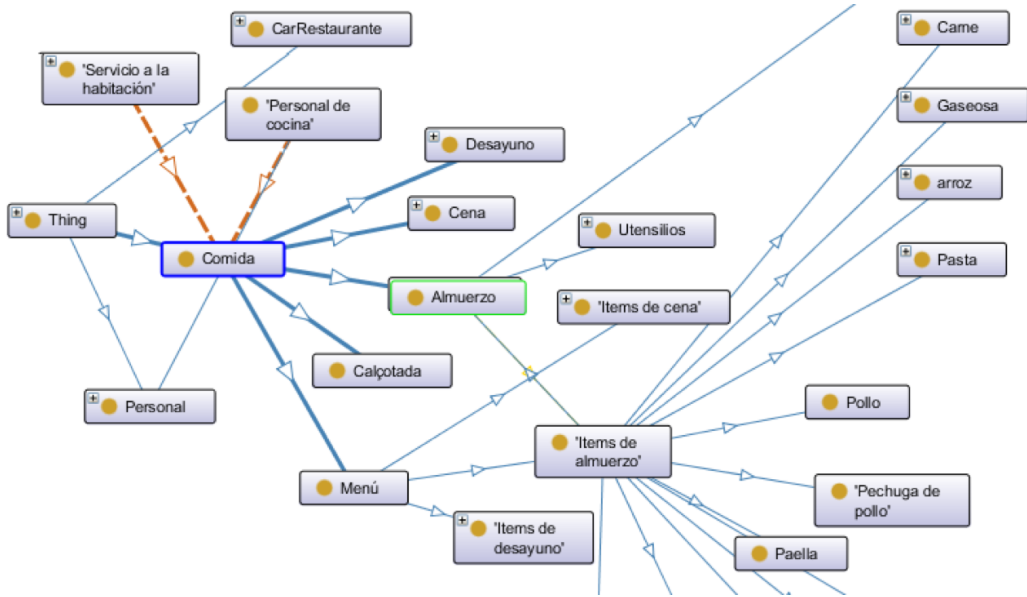


Figura 3: Partes de la ontología utilizada para análisis de opiniones de restaurantes

a la misma opinión. El *slot2* consiste en detectar la “Expresión Destino de la Opinión” (Opinion Target Expression, OTE) de un par E#A, esto es, la expresión lingüística usada en la opinión para hacer referencia a la entidad (E) y al atributo (A). Pueden haber opiniones para los que la OTE sea nulo. Existe una tarea que agrupa el *slot1* y el *slot2* que consiste en detectar las categorías existentes en la opinión y asignarles su correspondiente OTE. Esta tarea se denomina *slot1,2*. En el *slot3* se debe determinar la polaridad (positiva, negativa, neutra) de cada OTE.

En la Figura 2 se muestra un ejemplo de anotación de una opinión tomada del conjunto de datos de entrenamiento.

En este trabajo hemos abordado las siguientes subtareas: la subtask que aborda los *slot1* y *slot2* de manera conjunta (*slot1,2*); la subtask correspondiente al *slot2*; y finalmente, la subtask correspondiente al *slot3*.

Para ello, se ha usado el corpus de la tarea que consta de 2070 frases de entrenamiento y de 881 frases de evaluación. Como medida de evaluación para los *slot1*, *slot2* y *slot1,2* fue utilizada la medida F_1 y para el *slot3* la medida que se utilizó fue *Accuracy*.

Para abordar las subtareas de los *slot1,2* y *slot2* se utilizó la ontología multilingüe “Hontology” correspondiente al dominio de restaurantes considerando sólo la parte en español. Además, esta ontología se extendió añadiendo aquellas instancias que aparecían en el conjunto de entrenamiento propor-

nado para las subtareas. La Figura 3 muestra parte de la ontología resultante. Como resultado de esta extensión, el número de clases de la ontología ha pasado de 284 a 314, el número de propiedades de los objetos de 8 a 12 y el número de individuos de 0 a 258.

Tarea	Sistema	SemEval2016
slot 2 (F_1)	73.07	GTI/C/68.51
slot 1,2 (F_1)	46.24	TGB/C/41.21

 Tabla 1: Resultados de nuestro sistema en la Task5-SB1 en español frente al mejor sistema de SemEval2016 para los *slot1,2* y *slot2*

Los resultados obtenidos por nuestro sistema junto a los mejores resultados de SemEval para las subtareas correspondientes a los *slot1,2* y *slot2* se muestran en la Tabla 1. Como se puede observar, nuestro sistema obtiene valores de F_1 superiores a los ganadores de la competición. En la competición de SemEval, los mejores resultados para el *slot1,2* los obtuvo el equipo TGB (Çetin et al., 2016) y para el *slot2* el mejor equipo fue GTI (Álvarez López et al., 2016).

Analizando los resultados de la extracción de aspectos (*slot2*), cabe destacar que la elección y utilización de la ontología de dominio resultó satisfactoria para la identificación de aspectos, ya que estas representan los conceptos de un determinado dominio y sus relaciones, es decir, son un modelo abstracto de un dominio, donde los conceptos utilizados

están claramente definidos y no son simples diccionarios. Para la subtask definida en el *slots1,2*, aunque el sistema no está construido para tal fin, presenta buenos resultados solo realizando una mapeado con la ontología.

Para el *slot 3* aprendimos un modelo basado en SVM con kernel lineal. En un primer experimento se utilizó únicamente el corpus de entrenamiento proporcionado en la competición. Se utilizaron como representación de las opiniones los coeficientes tf-idf de segmentos de caracteres de longitud 7 utilizando el concepto de bolsa de caracteres. Los parámetros fueron elegidos mediante un proceso de validación cruzada de 10 iteraciones (10-fold cross validation). En un segundo experimento se añadieron lexicones de polaridad. En concreto se utilizaron el diccionario ELHUYAR (Saralegi y San Vicente, 2013) lematizado y los lexicones SOL e iSOL (Molina-González et al., 2013). Para este modelo, se utilizaron como características secuencias de hasta 7 caracteres. A estas características se les añadió como nuevas características, el número de palabras positivas y negativas contenidas en los lexicones mencionados. El número total de características del modelo final fue de 111058. La Tabla 2 muestra los resultados de los dos experimentos realizados para el *slot3* junto al mejor resultado obtenido en la competición SemEval.

Sistema	Accuracy
Sin lexicones	83.21
Con lexicones	84.79
SemEval2016 (IIT-T./U)	83.58

Tabla 2: Resultados de nuestro sistema en la task5-SB1 en español frente al mejor sistema de SemEval2016 para el *slot3*

Como se puede ver el uso de lexicones mejora considerablemente los resultados, más de un punto y medio. Estos resultados consiguen superar los mejores resultados obtenidos en la competición SemEval2016 Task5-SB1 *slot3* por el equipo IIT-T (Kumar et al., 2016), obteniendo 84.79 % de *Accuracy* frente a 83.58 %.

5 Conclusiones y trabajos futuros

En este trabajo se ha presentado un sistema que usa ontologías y aprendizaje automático. Se han logrado resultados interesantes y prometedores que superan los obtenidos por

los participantes de la competición SemEval. Un 73.07 en F_1 en la extracción de aspectos (*slot2*) y un 46.24 de F_1 en la subtask correspondiente a *slot1,2* utilizando una aproximación basada en ontologías. Para la subtask de clasificación de sentimientos (*slot3*) se ha obtenido una *Accuracy* del 84.79 % utilizando una aproximación basada en Máquinas de Soporte Vectorial y lexicones de polaridad.

A la vista de los buenos resultados obtenidos, nos planteamos como trabajo futuro explorar nuevos mecanismos que nos permitan integrar la información de las ontologías en los algoritmos de aprendizaje automático y así poder abordar todas las tareas conjuntamente así como la extensión a otros idiomas y dominios cubiertos por la ontología.

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto ASLP-MULAN: Audio, Speech and Language Processing for Multimedia Analytics (MINECO TIN2014-54288-C4-3-R y fondos FEDER). La estancia realizada, de enero a marzo de 2017, por Carlos Henríquez en la UPV, ha sido subvencionada por el programa Colciencias (convocatoria 727), Universidad Nacional de Medellín y Universidad Autónoma del Caribe Barranquilla.

Bibliografía

- Álvarez López, T., J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, y F. J. González-Castaño. 2016. Gti at semeval-2016 task 5: Svm and crf for aspect detection and unsupervised aspect-based sentiment analysis. En *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, páginas 306–311, San Diego, California, June. Association for Computational Linguistics.
- Cadilhac, A., F. Benamara, y N. Aussenac-Gilles. 2010. Ontolexical resources for feature based opinion mining : a case-study. En *Proceedings of the 6th Workshop on Ontologies and Lexical Resources (OntoLex 2010)*, páginas 77–86.
- Çetin, F. S., E. Yıldırım, C. Özbey, y G. Eryiğit. 2016. Tgb at semeval-2016 task 5: Multi-lingual constraint system for aspect based sentiment analysis. En *Proceedings of the 10th International Workshop*

- on *Semantic Evaluation (SemEval-2016)*, páginas 337–341, San Diego, California, June. Association for Computational Linguistics.
- Chaves, M., L. Freitas, y R. Vieira. 2012. Hontology: a Multilingual Ontology for the Accommodation Sector in the Tourism Industry. En *CTIC/STI - Comunicações a Conferências*, páginas 149–154.
- De Freitas, L. A. y R. Vieira. 2013. Ontology-based Feature Level Opinion Mining for Portuguese Reviews. En *Proceedings of the 22nd International Conference on World Wide Web. ACM*,, páginas 367–370.
- González, J.-A., F. Pla, y L.-F. Hurtado. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. En *Proceedings of the 11th International Workshop on Semantic Evaluation (pendiente de publicación)*, SemEval '17, páginas 723–727, Vancouver, Canada, August. Association for Computational Linguistics.
- Gonzalez-Agirre, A. y G. Rigau. 2013. Construcción de una base de conocimiento léxico multilingüe de amplia cobertura: Multilingual Central Repository Building a wide coverage multilingual lexical knowledge base: Multilingual Central Repository. *Linguamatica*, 5(1):13–28.
- Henríquez, C. y J. Guzmán. 2016. Las ontologías para la detección automática de aspectos en el análisis de sentimientos. *Revista Prospectiva*, 14(2):90 – 98.
- Henríquez Miranda, C., J. Guzmán, y D. Salcedo. 2016. Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. *Procesamiento del lenguaje Natural*, 56:25–32.
- Hu, M. y B. Liu. 2004. Mining and Summarizing Customer Reviews. En *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*,, páginas 168–177.
- Hurtado, L.-F. y F. Pla. 2014. Análisis de Sentimientos, Detección de Tópicos y Análisis de Sentimientos de Aspectos en Twitter. En *TASS 2014*.
- Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Martínez-Cámara, E., y Ureña-López, L. A. 2016. Combining resources to improve unsupervised sentiment analysis at aspect-level. *Journal of Information Science*, 42(2):213–229.
- Kontopoulos, E., C. Berberidis, T. Dergiades, y N. Bassiliades. 2013. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, páginas 4065–4074.
- Kumar, A., S. Kohail, A. Kumar, A. Ekbal, y C. Biemann. 2016. Iit-tuda at semeval-2016 task 5: Beyond sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis. En *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, páginas 1129–1135, San Diego, California, June. Association for Computational Linguistics.
- Lecun, Y., Y. Bengio, y G. Hinton. 2015. Deep learning. *Nature*, 521(7553):436–444, 5.
- Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- Liu, B. 2015. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
- Liu, B., M. Hu, y J. Cheng. 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web. En *Proceedings of the 14th international conference on World Wide Web. ACM*, páginas 342–351.
- Manek, A., P. Shenoy, y M. Mohan. 2016. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World Wide Web*, páginas 1–20.
- Marcheggiani, D., O. Täckström, A. Esuli, y F. Sebastiani. 2014. Hierarchical multi-label conditional random fields for aspect-oriented opinion mining. En *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, páginas 273–285.
- Medhat, W., A. Hassan, y H. Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, páginas 1093–1113.

- Meng, L., R. Huang, y J. Gu. 2013. A Review of Semantic Similarity Measures in WordNet. *International Journal of Hybrid Information Technology*, 6(1):1–12.
- Mikolov, T., K. Chen, G. Corrado, y J. Dean. 2013a. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.
- Mikolov, T., I. Sutskever, K. Chen, G. Corrado, y J. Dean. 2013b. Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546.
- Molina-González, M. D., E. Martínez-Cámara, M.-T. Martín-Valdivia, y J. M. Perea-Ortega. 2013. Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications*, 40(18):7250–7257.
- Padró, L. y E. Stanilovsky. 2012. FreeLing 3.0: Towards wider multilinguality. En *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.
- Pang, B. y L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(12):1–135.
- Peñalver-Martínez, I., F. García-Sánchez, R. Valencia-García, M. Ángel Rodríguez-García, V. Moreno, A. Fraga, y J. L. Sánchez-Cervantes. 2014. Feature-based opinion mining through ontologies. *Expert Systems with Applications*, 41(13):5995–6008.
- Pla, F. y L.-F. Hurtado. 2014. Political tendency identification in twitter using sentiment analysis techniques. En *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, páginas 183–192, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.
- Pontiki, M., D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. María Jiménez-Zafra, y G. Eryigit. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. En *SemEval*, páginas 19–30.
- Poria, S., I. Chaturvedi, E. Cambria, y F. Bisoio. 2016. Sentic LDA: Improving on LDA with Semantic Similarity for Aspect-Based Sentiment Analysis. En *Neural Networks (IJCNN)*.
- Qiu, G., B. Liu, J. Bu, y C. Chen. 2011. Opinion Word Expansion and Target Extraction through Double Propagation. *Computational linguistics*, 37:9 – 27.
- Saralegi, X. y I. San Vicente. 2013. Elhuyar at tass 2013. En *Proceedings of the TASS workshop at SEPLN 2013*, páginas 143–150. IV Congreso Español de Informática.
- Steinberger, J., T. Brychcín, y M. Konkol. 2014. Aspect-Level Sentiment Analysis in Czech. En *Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, páginas 24–30.
- Studer, R., V. R. Benjamins, y D. Fensel. 1998. I DATA & KNOWLEDGE ENGINEERING. *Data & Knowledge Engineering*, 25:161–197.
- Wang, H., Y. Lu, y C. Zhai. 2010. Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach. En *KDD’10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM.
- Xianghua, F., L. Guo, G. Yanyan, y W. Zhiqiang. 2013. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon.
- Zhang, W., H. Xu, y W. Wan. 2012. Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications*.
- Zhou, Z.-H., J. Wu, y W. Tang. 2002. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 137(1):239 – 263.